

ORIGINAL ARTICLE



Sample size determination in bioequivalence studies using statistical assurance

A. Ring^{1,2} | B. Lang³ | C. Kazaroho⁴ | D. Labes⁵ | R. Schall^{1,6} | H. Schütz⁷

¹ University of the Free State, Bloemfontein, South Africa

² medac, Wedel, Germany

³ Boehringer Ingelheim, Biberach, Germany

⁴ AIMS Rwanda, Kigali, Rwanda

⁵ Consultant, Berlin, Germany

⁶ IQVIA Biostatistics, Bloemfontein, South Africa

⁷ BEBAC, Vienna, Austria

Correspondence

Arne Ring, medac GmbH, Theaterstr. 6, 22880. Wedel, Germany.

Email: ringa@ufs.ac.za

Aims: Bioequivalence (BE) trials aim to demonstrate that the 90% confidence interval of the T/R-ratio of the pharmacokinetic metrics between two formulations (test [T] and reference [R]) of a drug is fully included in the acceptance interval [0.80, 1.25]. Traditionally, the sample size of BE trials is based on a power calculation based on the intrasubject variability coefficient of variation (CV) and the T/R-ratio of the metrics. Since the exact value of the T/R-ratio is not known prior to the trial, it is often assumed that the difference between the treatments does not exceed 5%. Hence, uncertainty about the T/R-ratio is expressed by using a fixed value for the sample size calculation. We propose to characterise the uncertainty about the T/R-ratio by a (normal) distribution for the log(T/R-ratio), with an assumed mean of $\log \theta = 0.00$ (i.e. $\theta = 1.00$) and a standard deviation σ_u , which quantifies the uncertainty. Evaluating this distribution leads to the statistical assurance of the BE trial.

Methods: The assurance of a clinical trial can be derived by integrating the power over the distribution of the input parameters, in this case, the assumed distribution of the log(T/R)-ratio. Because it is an average power, the assurance can be interpreted as a measure of the probability of success that does not depend on a specific assumed value for the log(T/R)-ratio. The relationship between power and assurance will be analysed by comparing the numerical outcomes.

Results: Using the assurance concept, values of the standard deviation for the distribution of potential log(T/R)-ratios can be chosen to reflect the magnitude of uncertainty. For most practical cases (i.e. when $0.95 \leq \theta \leq 1.05$), the sample size is not, or only slightly, changed when $\sigma = |\log(\theta)|$.

Conclusion: The advantage of deriving the assurance for BE trials is that uncertainty is directly expressed as a parameter of variability.

KEYWORDS

bioequivalence, crossover trial, sample size determination, statistical power, trial design

1 | INTRODUCTION

1.1 | Background

Bioequivalence (BE) studies are performed to demonstrate the pharmacokinetic similarity of a new drug product (test T)

compared with an established product (reference R). These studies are typically conducted in healthy volunteers in specialised units for human pharmacology trials. Setup and statistical analysis of BE trials are highly regulated worldwide [EMA¹, FDA²], leaving only limited room for methodological changes or improvements.

The key objective of a BE trial is the comparison of the pharmacokinetic metrics C_{max} and AUC, which summarise the concentration–time profile obtained after a single administration of either the test or the reference product. Both parameters are assumed to be log-normally distributed,^{3,4} so that the comparison can be performed by estimating the ratio

$$\theta = \frac{M^T}{M^R} \quad (1)$$

where M^T and M^R are the geometric means of the individual metrics (C_{max} or AUC) for the test and reference product.

After logarithmic transformation, the analysis is performed using a mixed model with adjustments for appropriate covariates, and the treatment contrast $\log(\theta) = \log(M^T) - \log(M^R)$ and its 90% confidence interval are derived. BE is demonstrated, if the backtransformed 90% confidence interval is fully included within the BE margins, typically with a lower bound of 80% and an upper bound of 125% (Figure 1). This procedure is called two–one-sided test procedure (TOST),⁵ as it is based on hypothesis testing with one-sided tests with $\alpha = 0.05$ as the level of significance.

Crossover designs—both drugs are given to the trial subjects in random order—are often used due to their statistical efficiency because statistical inference is based on the within-subject variability while the variability between subjects is removed from the analysis. This investigation is based on the 2×2 crossover design with 2 formulations and 2 periods.

As is typically the case, we assume that the coefficient of variation (CV) of the maximum plasma concentration (C_{max}) is larger than that of the area under the plasma concentration–time curve (AUC). Due to the high correlation between AUC and C_{max} , we thus only consider one TOST using the CV of C_{max} for the determination of the sample size, but the methods could be extended directly to include both AUC and C_{max} simultaneously.

BE trials are generally confirmatory trials, as they aim to reject the null hypotheses of inequivalence for both metrics. The statistical power for a BE trial is the probability to correctly reject the null hypothesis when the alternative hypothesis of equivalence is true, which is derived conditional on fixed values of all input parameters, in particular the assumed values for $\log(T/R)$ -ratio and CV.

The key parameter for planning BE trials is the sample size, as it will determine the cost and duration of the trial, and often practical considerations about subjects' recruitment schedules. Hence, a proper sample size calculation is crucial to ensure that the trial objectives can be met with sufficient probability at appropriate costs.

1.2 | Traditional sample size determination

Six parameters enter the sample size calculation for a BE trial⁶:

- Significance level $\alpha = 0.05$ one-sided – the error that equivalence is shown despite inequivalent formulations

What is already known about this subject

- Sample size planning for bioequivalence trials needs assumptions on the relative bioavailability (BA) of tested products and the variability of the pharmacokinetic metrics.
- Some trial sponsors assume 1.00 as test–reference ratio, others use 5% or similar deviations from 1.00 to account for potential differences of the formulations.
- In efficacy trials, the assurance concept is used for planning the sample size under uncertainty.

What this study adds

- The statistical assurance using a distribution of the test–reference ratio is a useful concept for planning bioequivalence trials.
- The uncertainty parameter is easier to interpret than a fixed deviation from 1.00, while it has similar statistical properties.

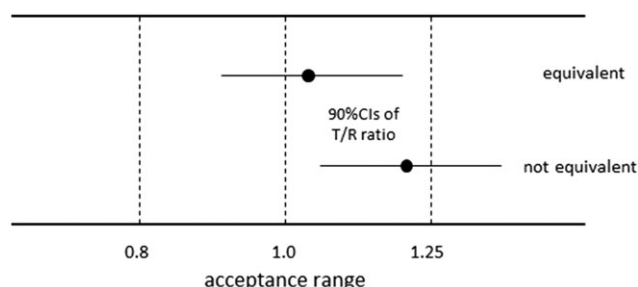


FIGURE 1 Acceptance range for bioequivalence tests. CI, confidence interval

- Type-II-error β – the error that equivalence cannot be shown despite equivalent formulations; the value $\pi = 1 - \beta$ is the power of the trial,
- BE margins $m_1 = 0.8$ and $m_2 = \frac{1}{m_1} = 1.25$,
- Expected ratio of the BE metrics $\theta = \frac{M^T}{M^R}$,
- Intraindividual coefficient of variation CV,
- Total sample size N .

The significance level and the margins are fixed by regulatory guidelines as indicated above. The coefficient of variation can often be taken from previous pharmacokinetic trials or literature data of the same drug, as the physicochemical drug properties may control most of the pharmacokinetic variability of the drug products. The power is generally chosen as a fixed value between 80 and 90%, depending on the sponsor's preferences for the desired probability to meet the trial objective.

The focus of this investigation is the expected T/R-ratio θ .

1.3 | Sample size determination of BE studies in the literature

When examining published reports of BE trials, it is apparent that there are different approaches to select the expected T/R-ratio for the sample size determination. In a systematic review of recent reports (2013–2018) of standard crossover BE trials⁷ it was found that out of 48 reports that described the details of the sample size calculation, 12 (25%) trials planned with an expected T/R-ratio of $\theta = 1.0$, hence they assumed the maximum power (e.g. Radicioni et al⁸, Bosilkovska et al⁹). In 16 (33%) trials, a value different from 1.0 was used—in the majority of cases either 0.95 or 1.05 (e.g. Ermer et al¹⁰). The remaining 20 (42%) reports used the limits of a range, such as 0.95–1.05 (e.g. Luo et al¹¹).

This demonstrates a certain heterogeneity of approaches to the planning of BE trials. Notably, no or only insufficient detail of the sample size calculation was provided in 78 trial reports, indicating that this information is often neglected despite the confirmatory nature of such trials.

1.4 | Objective

The aim of this investigation is to derive the probability of a successful BE trial not only for a single value of the T/R-ratio θ , but for a potential range of ratios. The concept used is called *statistical assurance*,^{12,13} which has been introduced for superiority trials. This concept shall be used to determine the required sample size and to compare outcomes with conventional power calculations.

2 | METHODS

2.1 | Power of a BE trial

In a sample size determination using the power approach, the power is determined for exactly one value of each of the 5 other parameters, of which two are fixed by the regulatory guidelines (significance level α and margins $m_1 = \frac{1}{m_2}$), as stated in Section 1.2.

Statistically, the power is determined *conditional* on these parameters, in particular on the expected ratio θ and the coefficient of variation CV.

Typically, the parameter about which the least information is available when determining a sample size is the ratio θ , because BE trials are often performed as *one shot* trials without pilot evaluations in humans. The *best guess* for this ratio is often 1.00, as there is generally no reason to think that one drug formulation has higher bioavailability (C_{\max} or AUC) than the other.

By contrast, if there was a difference between the two formulations, this would lead to a loss of power of the BE trial. Figure 2 shows the power curves for different values of the ratio θ . The power has a maximum at $\theta = 1.0$, because the distance to both margins is at its maximum (on the logarithmic scale). Any deviation from unity leads to a loss of power.

Various formulas to derive the power are available in the literature, most of which are approximate (but often sufficiently precise⁶). An exact solution was provided by Phillips,¹⁴ based on Owen's Q-function. This derivation is reproduced in the appendix; it has been implemented in the R-package PowerTOST.¹⁵

In the following, we denote the power for the BE trial by $\pi(\theta, CV, n)$. When we fix the sample size N and the coefficient of variation CV, we use the abbreviated symbol $\pi(\theta)$ for the power as a function of the T/R-ratio θ .

2.2 | Statistical assurance of a BE trial

We determine the probability of successfully rejecting the null hypothesis of inequivalence for a potential *distribution* of T/R-ratios. This distribution is formed based on uncertainty about the T/R-ratio at the time of planning the trial. Thus, instead of taking a single value of θ , a *distribution* Θ of potential ratios θ is considered for the assurance. The statistical assurance γ is then derived as the expected value of the power over the values of θ , namely:

$$\begin{aligned}\gamma &= E(\pi(\theta)) \\ &= \int \text{Power}(\eta) \cdot \text{Weight}(\eta) d\eta.\end{aligned}\quad (1)$$

The assurance concept is illustrated in Figure 3. The concept follows the introduction of Chuang-Stein¹² applied to the context of BE.

Figure 3 shows three panels for the components of the assurance. Panel (A) is the power function for varying values of the T/R ratio θ ,

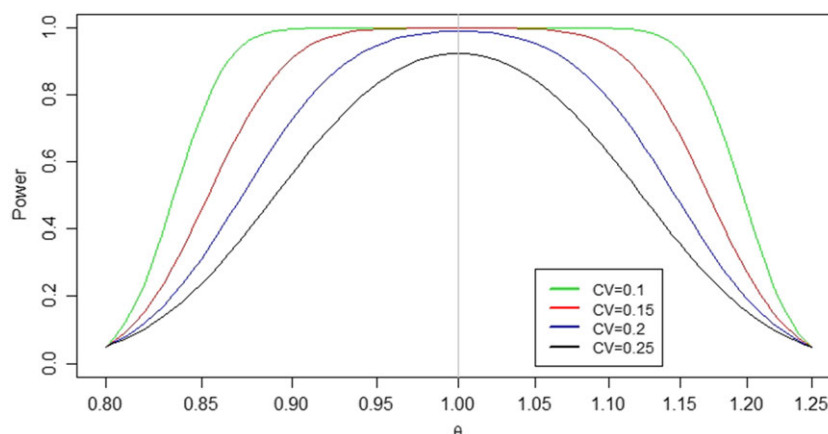


FIGURE 2 Power of a 2×2 crossover trial using 30 subjects, as a function of the T/R-ratio θ , for various values of the coefficient of variation (CV)

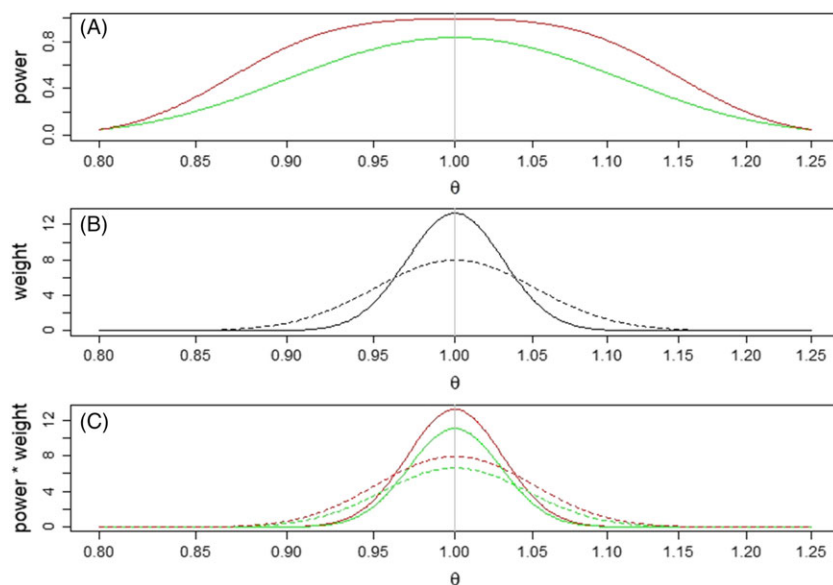


FIGURE 3 The components of the assurance concept. (A) The power curve (similar to Figure 2) for $n = 32$ (red) and $n = 16$ (green) subjects (coefficient of variation $CV = 0.2$). (B) Probability density functions for θ , assuming a normal distribution with a standard deviation σ_u of 0.03 (solid) and 0.05 (dashed). (C) The product of the functions in A and B, and all 4 combinations of N and σ_u are shown. The assurance is the area under this curve, which is 98, 80, 95 and 74%

which is similar to the graphs of Figure 2. It shows that the maximum power is achieved for $\theta = 1$. For deviations from unity, the power declines towards both margins.

Figure 3B shows the distribution function Θ , which expresses the probability for each specific value of θ . We have selected a distribution function which takes its largest value at $\theta = 1.00$. This means that the probability for a T/R-ratio that is in the centre is at its maximum, because in the absence of conflicting information (e.g. different dissolution data) it is likely that the BE metrics following administration of the two formulations are similar.

However, it is also reasonable to assume that the true value of the T/R-ratio differs slightly from 1, for example because of inherent variability of the production process of the formulation. Both directions would be equally likely, so the distribution function should be symmetric. An appropriate candidate for the weight function is the normal distribution $N(0, \sigma_u)$ with mean 0 and standard deviation σ_u . The σ_u is the *uncertainty parameter* for the distribution of $\log(\Theta)$. Less uncertainty implies more probability around the centre value, while more uncertainty (with a larger σ_u) leads to larger probabilities for further deviation from the optimum T/R-ratio.

The third panel (Figure 3C) shows the product of the above values for each value of θ . As introduced in (1) the assurance γ is the area under this product curve, which can be specifically written as follows:

$$\gamma(\sigma_u, CV, N) = \int \pi(\eta, CV, N) W_{\sigma_u}(\eta) d\eta \quad (2)$$

here $\pi(\eta, CV, n)$ is the power function (using a fixed $\alpha = 0.05$) and $W_{\sigma_u}(\eta)$ is the density function of the normal distribution (with mean 0 and standard deviation σ_u). There is no closed formula for the determination of the assurance, so that the integral (2) has to be derived numerically.

It should be noted that the limiting case when $\sigma_u \rightarrow 0$ corresponds to *no uncertainty* of the ratio θ , simply leads to the power (with ratio $\theta = 1$):

$$\gamma(0, CV, N) = \pi(1, CV, N) \quad (3)$$

In the following, the statistical assurance of BE trials will be examined based on its relationship to the underlying parameters, and compared to the power approach.

One of the key parameters for determining the conventional power π is the T/R-ratio θ . It will be shown that the new uncertainty parameter σ_u affects the statistical assurance γ in a similar way. Furthermore, the application of both methods will be illustrated by comparing the resulting sample sizes.

The relevant simulations were performed using the statistical software R,¹⁶ in particular, using the package PowerTOST.¹⁵

3 | RESULTS

3.1 | Qualitative comparison of power and assurance

The relationship between power/assurance and the sample size is shown in Figure 4 based on 2 values of the coefficient of variation CV. Each panel shows various graphs that indicate the power/assurance based on different values of the ratio θ or of the uncertainty parameter σ_u .

The qualitative relationship between sample size and assurance for different values of σ_u (Figure 4) is similar to the relationship between power and different values of θ . The assurance increases with sample size and decreases with the increasing values of the uncertainty parameter σ_u .

However, the curves differ somewhat when larger deviations of θ from the value 1.00 are assumed, compared to larger values of σ_u . The power curve increases slowly but steadily with increasing sample size, while the assurance increases more rapidly in the low sample size range, but then flattens somewhat, so that more assurance requires a much higher sample size. The reason for this observation is that with larger values of σ_u , the tail of the weight function (when θ comes close

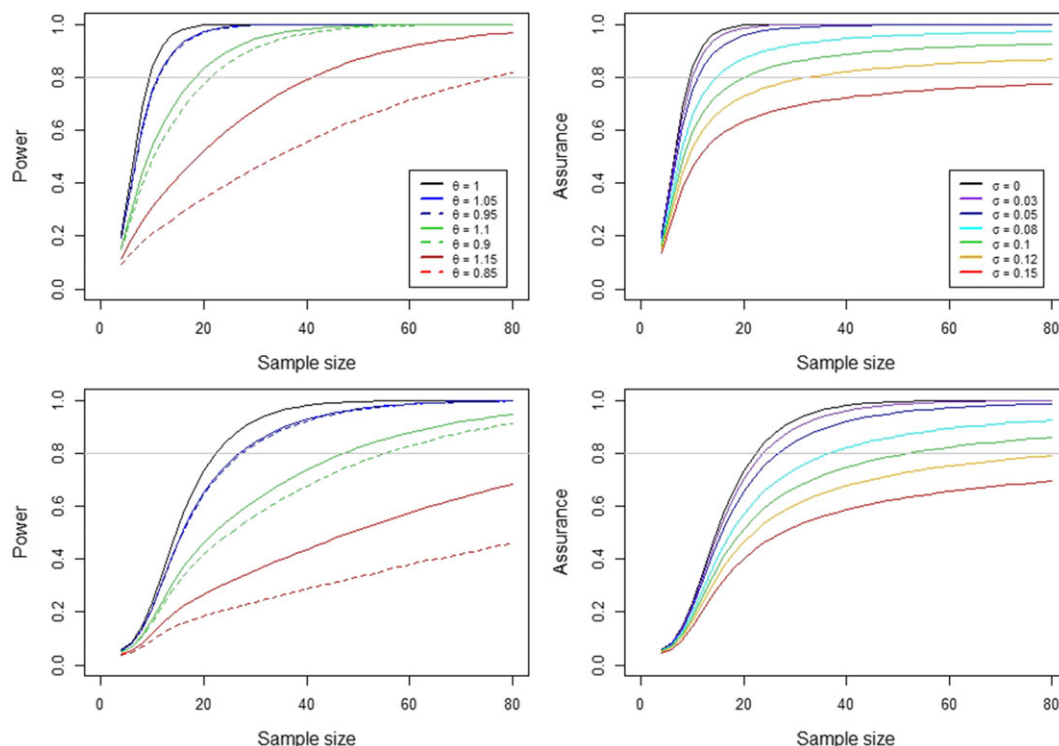


FIGURE 4 Power (left) and assurance (right) of a 2×2 crossover trial for a fixed coefficient of variation (CV) of 15% (above) and 25% (below) as a function of sample size, for different values of θ (left) and σ_u (right) (while $\theta = 1$). The colours are the same for $\theta = 1 \pm \sigma_u$, while additional intermediate values of σ_u are shown for the assurance graphs

to the BE margins) becomes heavier, while, by contrast, the power is indeed very low for these values of θ , so that the assurance remains low. This is much less the case for $\sigma_u \leq 0.05$, because the tail of the weight function is less heavy close to the margin.

Figure 5 compares the sample size required to reach predefined levels of power (as a function of the expected T/R-ratio θ) and levels of the assurance (as a function of the uncertainty parameter σ_u).

For the power concept, the curve is U-shaped and hence symmetric, because the sample size has to be increased when $\log(\theta)$ deviates from 0.00 to either side. For the assurance curve, the sample size increases monotone with increasing values of σ_u because the uncertainty parameter σ_u can only be positive.

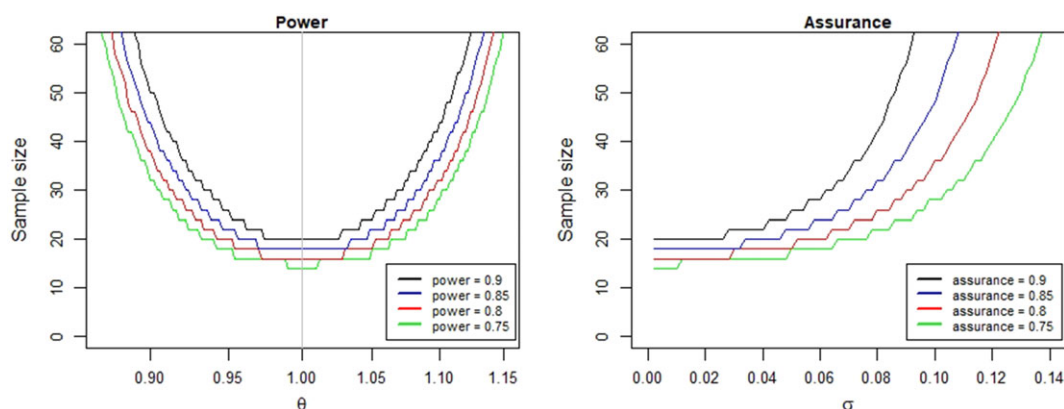


FIGURE 5 Required sample size for a 2×2 crossover bioequivalence trial as a function of the expected T/R-ratio θ (left) or of the uncertainty σ_u (right), for different values of the desired power (left) /assurance (right)

For positive values of θ , the qualitative behaviour of the assurance curves is similar to that of the power curves. When comparing the required sample sizes in case of $\sigma_u = |\log(\theta)|$, slightly more sample size is required to achieve the same level of assurance compared with the power.

3.2 | Quantitative comparison

Differences between the assurance and power concepts shall be illustrated in a numerical example. A BE trial is planned for a new formulation Temozolomide, based on data from a previous trial.¹⁷ This trial has shown an intraindividual CV of 21.4% for C_{\max} and 5.2% for

AUC. The sponsor of the new trial would like to achieve a probability of success of 90%.

The standard sample size calculation using a power calculation with $\theta = 1.00$ leads to a sample size of 22 subjects. If the sponsor wants to account for possible deviations using a different θ , the sample size would slightly increase, e.g. for $\theta = 0.95$ the required sample size would be 28 subjects (Table 1a).

Accounting for the possible uncertainty of the T/R-ratio using the assurance concept, the sample size to achieve a statistical assurance of 90% with the uncertainty parameter $\sigma_u = 0.05$ is also 28 subjects (Table 1b).

Table 1 shows that the sample sizes mostly are similar when using power and assurance. However, the sample size would be lower for an assurance of 80% with $\sigma_u = 0.06$ compared to a power of 80% with $\theta = 0.94$. By contrast, the sample sizes are increased for the assurance concept when the uncertainty σ_u gets larger.

4 | DISCUSSION

4.1 | General comparison between power and assurance

The traditional sample size calculation for BE trials is based on several parameters, including the expected T/R-ratio θ . The statistical power for the BE trial is determined using fixed values for these parameters.

The test formulation has typically not been administered to humans before, or only in small trials so that there is uncertainty about the pharmacokinetic concentration–time profile in humans. The optimal case is a T/R-ratio θ equal to 1.00. However, any deviation of θ from 1.00 would result in a loss of power, so that it is conservative to allow for deviations from the optimal value. Our systematic review has shown

TABLE 1 Sample sizes required to achieve (a) a *target power* for selected values of the anticipated T/R-ratio θ and (b) a *target assurance* for selected values of the uncertainty parameter σ_u . The targets have been set to 80 and 90%, the coefficient of variation (CV) was set to 21.4%. For the values in bold, the sample sizes differ between power and assurance concept

| Target power/ assurance | (a) Power calculation | | | (b) Assurance calculation | | |
|----------------------------|-----------------------|-----------|----------------|---------------------------|-----------|--------------------|
| | θ | n | Achieved power | σ_u | n | Achieved assurance |
| 80% | 1.00 | 18 | 83.3% | 0.00 | 18 | 83.3% |
| 80% | 0.95 | 22 | 82.4% | 0.05 | 22 | 83.3% |
| 80% | 0.94 | 24 | 81.7% | 0.06 | 22 | 80.0% |
| 80% | 0.93 | 26 | 80.1% | 0.07 | 26 | 81.9% |
| 80% | 0.92 | 28 | 80.2% | 0.08 | 30 | 80.3% |
| 90% | 1.00 | 22 | 91.6% | 0.00 | 22 | 91.6% |
| 90% | 0.95 | 28 | 90.4% | 0.05 | 28 | 90.4% |
| 90% | 0.94 | 32 | 90.9% | 0.06 | 32 | 90.3% |
| 90% | 0.93 | 36 | 90.5% | 0.07 | 38 | 90.2% |
| 90% | 0.92 | 42 | 90.8% | 0.08 | 48 | 90.2% |

that some trial sponsors follow such a conservative approach, while others assume a value of 1.00 for the T/R-ratio.

In this paper, we propose a methodology to account for the uncertainty in the T/R-ratio when determining the probability of a successful BE trial. Instead of assuming a *fixed* value different from 1.00, we assume a *distribution* of values (Θ). Specifically, we propose using a normal distribution (on log-scale) with the mean of 0.00 and the standard deviation σ_u (implying a geometric mean of 1 for the T/R-ratio, θ). The standard deviation σ_u is a measure of the uncertainty that the sponsor has about the T/R ratio.

The geometric mean of 1.00 for θ implies that we still assume that this value is the most likely for the T/R-ratio, while the distribution indicates that deviations from 1 are accommodated. There would be other options to select a distribution function Θ with a different shape. However, the normal distribution plays an important role in both Bayesian and frequentist statistics to model uncertainties or quantitative errors of mean values. Its simplicity, having a single parameter σ_u , motivates it as a good choice for modelling the T/R-ratio. Furthermore, pharmacokinetic metrics are generally log-normally distributed.⁴

As of now, the uncertainty parameter σ_u is a new parameter in the sample size consideration, and hence its value must be chosen empirically. However, this is not much different from the situation before when the value for the expected T/R-ratio had to be chosen empirically. For the determination of the assurance, the uncertainty parameter takes over the conservative fixed choice of T/R-ratio when it was selected different from 1.00—hence the number of parameters that are actually used does not increase.

One option for investigating the uncertainty parameter would be a systematic review of BE trials. As a preliminary result we found in our review⁷ that the variability of the estimated pharmacokinetic metrics is between 0.05 and 0.07.

The investigations in this paper have shown that for the range of θ between 0.95 and 1.05, the power curves are numerically similar to the assurance curves when $\sigma_u = |1 - \theta|$ or $\sigma_u = |\log(\theta)|$. Although the power and assurance curves are similar within this range, we have also seen that the assurance is always slightly smaller than the power, i.e. $\gamma(\sigma_u = |\log(\theta)|) < \pi(\theta)$. The numerical difference between the two methods becomes larger for larger deviations of θ from unity.

Despite this similarity, the two parameters are conceptually quite different: when a fixed value of $\theta \neq 1$ is used for the sample size determination using the power concept, this choice could imply that the sponsor believes that e.g. the C_{\max} of the test formulation is different to that of the reference. The assurance concept does not lead to such an interpretation because the assumed distribution is still centred around 1.00. This assurance concept just acknowledges that the sponsor is not sure that the pharmacokinetic metrics are exactly equal.

The difference between the power and the assurance concept becomes clear when σ_u is assumed to be much larger than 0.05: In this case, the distribution of T/R-ratios that the sponsor would accept as possible values would include values that are beyond the regulatory margins, and this should not be a realistic assumption. (This might actually be a more realistic explanation of the uncertainty of Θ than just using a fixed value of e.g. $\theta = 1.05$ in the sample size determination of a BE trial.)

4.2 | Implications

The use of the assurance concept for planning BE studies could change the communication about the underlying assumptions compared to the power concept. The standard deviation of the weight function, σ_u , can be understood directly as a measure of uncertainty, and pharmaceutical companies can select their level of uncertainty as a strategic decision for their planning of BE trials. Assuming a distribution of T/R-ratios instead of a single fixed value can be understood much better by nonstatistical scientists and physicians, in particular when the distribution around $\log(1.00)$ is symmetric. By contrast, the magnitudes of the parameter values for σ_u and θ are quite similar, so that the use of the assurance concept does not necessarily increase the sample sizes, at least in the relevant range of up to 5% difference.

The assurance concept appears to be of particular practical relevance because the specification of pharmaceutical drug formulations always leaves room for slight deviations: for example, typical regulatory limits for *batch-to-batch variabilities* of drug products are in the range of 5%. This variability is typically not covered by the drug-specific CV, because most clinical pharmacology studies which analyse the pharmacokinetic properties are performed with a single product batch.¹⁸ This might be another good reason for incorporating deviation from the optimal T/R-ratio into the sample size calculation of a BE trial.

Currently, it remains an open question whether the choice of $\sigma_u = 0.05$ could serve as a *best guess* for quantifying the uncertainty. The systematic review of recently published BE trials led to the preliminary finding that the distribution of the *estimated* θ for AUC and C_{max} for trials that successfully demonstrated BE has an estimated standard deviation of about 0.07. However, a more thorough review might be warranted, which should take into account the potential for publication bias affecting unsuccessful or underpowered BE trials.

4.3 | Potential extensions of the concept

The concept of statistical assurance has previously been introduced and discussed for superiority efficacy trials^{12,19} as the *probability of a successful trial*. In contrast, the power is the *probability of rejecting the null hypothesis when it is false*, which is derived conditional on fixed values of all input parameters.

It was found for those trials that typical sample sizes of phase III trials would lead to an assurance of between 70 and 80%, even if these trials had been powered for up to 90%. Hence, most trials still perform traditional power calculations, maybe because scientists are used to the concept. In this paper, we have shown that for BE trials the magnitude of power and assurance are similar for given sample size, so the concepts lead to similar sample sizes in the context of equivalence tests. Sponsors could still discuss which value for the uncertainty parameter they consider as most relevant for their trial planning.

The proposed concept for BE trials can be extended in several ways: first, it could also capture the uncertainty on the drug-specific coefficient of variation CV. As indicated above, this parameter has typically been estimated in previous pharmacokinetic trials of the drug,

but these estimates are also subject to uncertainty, which can be incorporated independently using a suitable distribution.²⁰ The uncertainty could be quantified directly in a meta-analysis of the available pharmacokinetic trials.

Secondly, the assurance concept could account for sample size determination when small differences in the dissolution or absorption would be expected, e.g. from outcomes of in vitro tests and modelled in vitro in vivo correlations. In this case, a $\theta \neq 1.00$ for the normal distribution could be incorporated in addition to the uncertainty parameter σ_u , and both parameters could be adjusted independently of each other, given the available information.

Overall the assurance concept for planning BE trials could be a valuable method to derive and explain sample size considerations. Sponsors should be encouraged to account for the uncertainty of the T/R-ratio, and the use of a particular parameter (σ_u) to account for the uncertainty seems to be a natural option that does not lead to substantially different sample sizes.

ACKNOWLEDGEMENTS

Parts of this investigation were performed by C. Kazaroho as her MSc project for the African Institute for Mathematical Sciences (AIMS) in Rwanda.²¹

COMPETING INTERESTS

There are no competing interests to declare.

CONTRIBUTORS

The study concept and design was developed by A.R. and C.K. All authors were responsible for the statistical analysis and participated and interpretation of data. C.K. and A.R. visualized the results. A.R. and B.L. drafted the manuscript. All authors revised the manuscript critically for important intellectual content, and approved the final version to be published.

DATA SHARING STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

ORCID

A. Ring  <https://orcid.org/0000-0002-4324-5820>
 C. Kazaroho  <https://orcid.org/0000-0003-3665-8974>
 D. Labes  <https://orcid.org/0000-0003-2169-426X>
 R. Schall  <https://orcid.org/0000-0002-4145-3685>
 H. Schütz  <https://orcid.org/0000-0002-1167-7880>

REFERENCES

1. European Medicines Agency, Guideline on the investigation of bioequivalence. 2010. Doc. Ref. CPMP/EWP/QWP/1401/98 Rev. 1.
2. Food and Drug Administration. 2014. Bioavailability and Bioequivalence Studies Submitted in NDAs or INDs—General Considerations. Draft guidance.

3. Lacey LF, Keene ON, Pritchard JF, Bye A. Common noncompartmental pharmacokinetic variables: are they normally or log-normally distributed? *J Biopharm Stat.* 1997;7(1):171-178.
4. Julious SA, Debnarot CA. Why are pharmacokinetic data summarized by arithmetic means? *J Biopharm Stat.* 2000;10(1):55-71.
5. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Br J Clin Pharmacol.* 2019;85:1622. P79.
6. Julious SA. Sample sizes for clinical trials with normal data. *Stat Med.* 2004;23(12):1921-1986.
7. Ring A, Kazaroho C, Awounvo SJ. Sample Size Planning in Bioequivalence Trials: A Systematic Review. *Br J Clin Pharmacol.* 2019;85:1622 P79.
8. Radicioni M, Castiglioni C, Giori A, Cupone I, Frangione V, Rovati S. Bioequivalence study of a new sildenafil 100 mg orodispersible film compared to the conventional film-coated 100 mg tablet administered to healthy male volunteers. *Drug des Devel Ther.* 2017;11:1183-1192.
9. Bosilkovska M, Samer C, Deglon J, et al. Evaluation of mutual drug-drug interaction within Geneva cocktail for cytochrome p450 phenotyping using innovative dried blood sampling method. *Basic Clin Pharmacol Toxicol.* 2016;119(3):284-290.
10. Ermer J, Corcoran M, Lassester K, Martin PT. Relative bioavailabilities of lisdexamfetamine dimesylate and d-amphetamine in healthy adults in an open-label, randomized, crossover study after mixing lisdexamfetamine dimesylate with food or drink. *Ther Drug Monitor.* 2016;38(6):769-776.
11. Luo Z, Nan F, Miao J, Chen Z, Li M, Liang M. Pharmacokinetics and bioequivalence of two formulations of febuxostat 40-mg and 80-mg tablets: a randomized, open-label, 4-way crossover study in healthy Chinese male volunteers. *PLoS One.* 2016;11(3):e0150661.
12. Chuang-Stein C. Sample size and the probability of a successful trial. *Pharm Stat.* 2006;5(4):305-309.
13. O'Hagan A, Stevens JW, Campbell MJ. Assurance in clinical trial design. *Pharm Stat.* 2005;4(3):187-201.
14. Phillips KF. Power of the two one-sided tests procedure in bioequivalence. *J Pharmacokin Biopharm.* 1990;18(2):137-144.
15. Labes D, Schütz H, Lang B. (2018) PowerTOST: Power and Sample Size Based on Two One-Sided t-Tests (TOST) for (Bio)Equivalence Studies. Ver 1.4-7. <https://cran.r-project.org/package=PowerTOST>.
16. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
17. Muggeri A, Vago M, Pérez S, et al. A randomized, open-label, two-way crossover, single-dose bioequivalence study of Temozolomide 200 mg/m² (Dralitem® vs. Temodal® capsules) in patients with primary tumors of the central nervous system under fasting conditions. *Drugs R D.* 2017;17(3):427-434. <https://doi.org/10.1007/s40268-017-0199-3>
18. Burmeister Getz E, Carroll KJ, Jones B, Benet LZ. Batch-to-batch pharmacokinetic variability confounds current bioequivalence regulations: a dry powder inhaler randomized clinical trial. *Clin Pharmacol Ther.* 2016 Sep;100(3):223-231. <https://doi.org/10.1002/cpt.373>
19. Chen DG, Ho S. From statistical power to statistical assurance: It's time for a paradigm change in clinical trial design. *Comm Statist - Simul Comput.* 2017;46(10):7957-7971.
20. Patterson S, Jones B. Bioequivalence and Statistics. In: *Clinical Pharmacology*. 2nd ed. Boca Raton: Chapman & Hall/CRC Press; 2017 Chapter 5.7 "Determining Trial Size".
21. Kazaroho C. Comparison of Sample Size Determination for Bioequivalence Trials using either Deterministic Power or Expected Power (Assurance Concept). MSc thesis, African Institute for Mathematical Sciences (AIMS), Rwanda. 2018.
22. Owen DB. A special case of a bivariate non-central t-distribution. *Biometrika.* 1965;52(3-4):437-446.

How to cite this article: Ring A, Lang B, Kazaroho C, Labes D, Schall R, Schütz H. Sample size determination in bioequivalence studies using statistical assurance. *Br J Clin Pharmacol.* 2019;85:2369-2377. <https://doi.org/10.1111/bcp.14055>

APPENDIX

Power of the TOST procedure

The power of a statistical test is the probability that the hypothesis H_0 (in our case bioinequivalence) is rejected if the alternative hypothesis H_1 (here bioequivalence) is true. In other words, the probability of correctly accepting bioequivalence is the power of the test.

The power of the two-one-sided t-tests (TOST) is thus given by

$$\text{Power} = \text{Prob}(t_1 \geq t_{(1-\alpha, n-2)} \text{ and } t_2 \leq -t_{(1-\alpha, n-2)} | \text{bioequivalence holds}) \quad (\text{I})$$

The t_1 and t_2 values are the t-test statistics of the TOSTs described above.

Owen²² has shown that the pair (t_1, t_2) has a specific bivariate non-central t-distribution and that the power based on that distribution can be calculated as the difference of two definite integrals (Owen's Q function):

$$\text{Power} = 1 - \beta = Q_{df}(-t_{(1-\alpha, df)}, \delta_2; 0, R) - Q_{df}(-t_{(1-\alpha, df)}, \delta_1; 0, R) \quad (\text{II})$$

where $t_{(1-\alpha, df)}$ is the $(1-\alpha)$ quantile of a t-distribution with df degrees of freedom. df is $(N-2)$ in case of a classical 2×2 cross-over design (using sample size N) and

$$\theta = \text{null ('true')} \text{ ratio} \quad (\text{IIa})$$

$$\delta_1 = \frac{\log(\theta) - \log(m_1)}{s_e \sqrt{2/N}}$$

$$\delta_2 = \frac{\log(\theta) - \log(m_2)}{s_e \sqrt{2/N}}$$

$$R = \frac{\sqrt{df}(\delta_1 - \delta_2)}{2 \cdot t_{(1-\alpha, df)}}$$

for log-transformed pharmacokinetic metrics, where s_e is the residual standard error, m_1 and m_2 are the lower and upper bioequivalence acceptance bounds (usually 0.8 and 1.25).

The residual variance (s_e^2) is connected to the within-subject coefficient of variation CV by

$$s_e^2 = \text{mse} = \log(CV^2 + 1)$$

$$CV = \sqrt{\exp(s_e^2) - 1}$$

Owen's Q-function is defined as:

$$Q_v(t, \delta; a, b) = \frac{\sqrt{2\pi}}{\Gamma\left(\frac{v}{2}\right) \cdot 2^{\frac{(v-2)}{2}}} \int_a^b \Phi\left(\frac{t \cdot x}{\sqrt{v}} - \delta\right) \cdot x^{v-1} \cdot \varphi(x) \cdot dx \quad (\text{III})$$

where $\Gamma(x)$ is the gamma-function, $\varphi(x)$ and $\Phi(X)$ are the density and cumulative distribution function of the standard normal distribution, respectively.

This exact algorithm is implemented in the R package *PowerTOST*¹⁵ via numerical evaluation of the definite integral using the `integrate()` function of the R package *stats*, part of the base R-project installation.